



Using Diagnostic Classroom Assessment: One Question at a Time

by Joseph F. Ciofalo & E. Caroline Wylie — January 10, 2006

The No Child Left Behind legislation requires states and districts to pay more attention to issues of testing and to concerns about adequate yearly progress. One positive outcome has been continued attention to the alignment of standards, curriculum, and assessment at the state, district, and school level. However, while a focus on testing can illuminate potential learning issues, testing alone cannot move learning forward: To improve student performance, classroom instruction needs to improve. In this article, the focus is on what happens in classrooms on a daily basis between teachers and students in particular, giving teachers a framework to make adjustments that can improve student learning. The article presents a brief discussion of formative assessment and locates the use of diagnostic items in teachers' daily practice as a specific formative assessment tool. This article is an analytic essay that focuses on the conceptual work of formative assessment that is the basis of a project in which 4th and 8th grade mathematics and science teachers are incorporating diagnostic items into their every day classroom practice. The article concludes with a suggestion that diagnostic items are not only a teaching resource that can be used to have immediate impact on classroom practice, but could also be used for professional development purposes.

INTRODUCTION

Increasingly, conversations about education have become conversations about accountability. The No Child Left Behind legislation requires states and districts to pay more attention to issues of testing and to concerns about adequate yearly progress. One positive outcome has been continued attention to the alignment of standards, curriculum, and assessment at the state, district, and school levels. However, while a focus on testing can illuminate potential learning issues, testing alone cannot move learning forward: To improve student performance, classroom instruction needs to improve. In this article, the focus is on what happens in classrooms on a daily basis between teachers and students—in particular, giving teachers a framework to make adjustments that can improve student learning.

A research synthesis by Paul Black and Dylan Wiliam (1998) indicates that a focus on formative assessment provides such a framework for improvement. They define formative assessment as frequent, interactive assessments of student progress and understanding in which evidence of learning is evoked and then used as feedback to adjust teaching and learning. Black and Wiliam's extensive analysis illustrates that using the lens of formative assessment to pay attention to what happens in the daily interactions between students and teachers has a positive impact on learning, even when measured on standardized tests.

This commentary will concentrate on one important aspect of daily classroom assessment—questioning. In particular; it will attend closely to a specific type of questioning that teachers can use to find out what their students know and, more importantly, what they do not know, so that their misconceptions can be addressed. These kinds of questions are called *diagnostic questions* or, to borrow a term from the testing world, *diagnostic items*. The testing term *item* is used because this type of question bears a resemblance to multiple-choice items seen on standardized tests. However, as will be clear, these items are being used in a decidedly different manner—to gather ongoing evidence in order to positively impact classroom teaching and learning.

The commentary will begin with a brief discussion of formative assessment so that the formative nature of this approach is clear. Then, the discussion will move to how teachers can use diagnostic items in their day-to-day teaching.

FORMATIVE ASSESSMENT

This article uses the definition of formative assessment suggested by Black et al. (2003): "Formative assessment is a process, one in which information about learning is *evoked* and then *used* to modify the teaching and learning activities in which teachers and students are engaged. [Emphasis in the original]" (p. 122). This definition takes the idea of formative assessment beyond the "micro-summative" assessments of classroom tests and homework (p. 122). It broadens the sources of evidence and solidifies the notion of what should be done *with* the evidence. The sources from which evidence can be evoked do not exclude information garnered from formal assessments, but more important than the source of evidence is the idea that the information obtained impacts subsequent teaching and learning activities—and sooner, rather than later.

One of the main strategies of formative assessment is the area-questioning, and the use of diagnostic items is part of this larger strategy. What makes diagnostic items particularly formative is that an incorrect response to a diagnostic item not only provides information that a student does not clearly understand a particular topic; it also provides specific insight

into *what it is that the student does not understand*—in other words, the nature of his/her misconceptions. Hence use of the term *diagnostic*—the student's response to the question helps the teacher “diagnose” the learning problem. However, teachers need to be sure they are asking a question that will best uncover student misconceptions. For example, a teacher might say to his/her students:

Put the following numbers in order, smallest to largest.

0.1 0.931 0.67 0.32

Students who do this correctly may really understand the concept of place value in decimal numbers; or they may simply be ignoring the decimal point and sorting the numbers 1, 931, 67, and 32. A more diagnostic approach would be:

Select the largest number from the following list.

0.94 1.25 1.4

A student who gets the correct answer for the original question using an incorrect strategy will likely select 1.25 as the largest number. With this evidence, the teacher has clearer evidence of her students' misconceptions and can use it to improve learning.

USING DIAGNOSTIC ITEMS TO ELICIT MISCONCEPTIONS

Using diagnostic tests to identify misconceptions can be a powerful tool to assist teachers in the “evoking” stage of formative assessment, but administering such tests can be very time-consuming. In contrast, using individual diagnostic items can be efficient and flexible. These are single, multiple choice questions connected to a specific content standard or objective. They have one or more answer choices that are incorrect but related to common student misconceptions regarding that standard or objective.

Misconceptions are inaccurate, incomplete, or partial understandings that students might have. For instance, students have many misconceptions connected to fractions. Students may not understand that there is an infinite number of fractions since they most commonly work with halves, thirds, quarters, fifths, sixths, and eighths. Furthermore, since students are first introduced to fractions as parts of a whole, they often believe that a mixed number with a whole number has to be larger than an improper fraction, e.g., $2\frac{1}{2}$ is greater than $20/2$. This partial understanding might well have sufficed during the introductory stages of their work on fractions, but it is not appropriate for all fraction situations.

In science, misconceptions often develop because of how students interact with and observe the world from an early age. In addition, everyday language often differs from scientific usage. So, while mathematics misconceptions may sometimes develop from rules that are not fully generalizable, science misconceptions often develop from students' attempts to understand the world around them. For example, as students attain an understanding of magnetism, over-generalization might lead them to believe that all metals or all silver-colored materials are attracted to magnets. If a student has an incorrect understanding of a central concept or an understanding that only holds in certain situations, further learning is going to be impeded until the student can develop a more complete or accurate understanding.

USING DIAGNOSTIC ITEMS IN THE CLASSROOM

We are working with a group of teachers to construct a bank of 50 to 100 multiple-choice items for 4th and 8th grade mathematics and science. Each item is connected to a particular content standard or objective, and at least one or more of the incorrect answers are related to misconceptions that students often have regarding that particular learning goal. The item bank is intended to serve as a complementary resource for teachers; it is not a full curriculum covering all possible content areas. Nor is it a series of tests to be administered at four-week intervals. Rather, the expectation is that teachers will use one item at a time, within the flow of day-to-day lessons. Teachers will be able to search through the database of items to identify one or two that are relevant to a particular instructional topic they will be focusing on in an upcoming lesson.

Asking questions is a natural part of every teacher's practice, but in the rush of teaching, it is sometimes difficult to think of quality questions that allow a teacher to tailor future instruction to the specific learning needs of students. For that reason, a great deal of effort is being spent constructing items for the database so that teachers have access to a bank of questions that are connected to key learning goals and to potential student misconceptions. A single item is quick to “administer”—most take only a couple of minutes to complete. Based on students' responses to the item, the teacher can give immediate feedback to an individual student, a group of students, or the whole class. Or, she might use the information to structure the next lesson.

For example, a common 4th-grade learning objective is “Use place value to read, identify, and write whole numbers.” Students have misconceptions about what to do when writing a number that has no hundreds or tens values. An item that

addresses the learning objective and has distracters that link to specific misconceptions:

- Write two thousand sixty seven as a number.
- (a) 267
 - (b) 2067
 - (c) 200067
 - (d) 2000607

The student who answers (c) or (d) treats each number in the sentence separately by writing 2000 and appending the 67, or by writing both 2000 and 60 and appending the 7. A different incorrect approach might lead a student to write down the numbers, but to ignore all zeroes as place holders, represented by option (a). With this question, the teacher can determine which students do not understand that the correct answer is (b); but more powerfully, she also has a potential diagnosis of what students incorrectly think about place value. Armed with this information, she can adjust instruction as necessary.

There is flexibility both in terms of when in a lesson a diagnostic item might be used and how a teacher might use it. The most important thing to consider is that the teacher is evoking information she needs to inform subsequent instruction or to provide feedback to students. Thus, an item might be used at the start of a lesson to determine a baseline of the class' understanding of a concept. A review could then be abbreviated or extended depending on student understanding. Similarly, close to the end of a sequence of instruction, a teacher might want to know which students are still struggling with the central ideas. A teacher could also choose to use an item mid-way through a lesson for which she has planned two possible instructional avenues. If students struggle with one aspect, she might instruct in a certain way; but if they struggle with a different aspect, her instruction would look quite different. By learning how teachers actually use the items in their different contexts and for different purposes, it is possible to develop a wide range of classroom applications and guidance for their use.

WHAT MAKES A GOOD DIAGNOSTIC ITEM?

There are certain guidelines that test developers follow when writing items for summative, high-stakes tests. For instance, test developers must pay attention to the issues of validity and reliability (Messick, 1989; Allen & Yen, 2002). Furthermore, in the interest of fairness, in a multiple-choice test that is mostly comprised of items with four answer options, a test developer will not include an item that has five possible answers; no matter how good those five might be, since this change in format might be confusing for students. Similarly, multiple-choice items that are ambiguous or have two correct answers will never deliberately be included on a high-stakes test, since a student who identifies both correct answers will be confused about how to respond to the item.

When it comes to developing individual diagnostic items to be used in the classroom, validity is still a critical issue. If a teacher wants to find out what students know about a particular topic, the item clearly must relate in a relevant and age-appropriate way to that topic. However, other constraints may be loosened in several interesting ways. For example, reliability becomes less of an issue because of the nature of the usage of the item. The teacher is not using the item to develop a score or grade from students' responses but rather to gain clearer insight into their thinking. The stakes are very low. Any decisions she makes that might not be appropriate because of the unreliability of a single item can be immediately corrected.

Moving away from a high-stakes environment also frees item writers from some other constraints. Since items will be administered individually, they could have two, three, four, five or more distracter options—whatever is most appropriate for a particular item. In fact, for some items it might be valuable to have two correct answers, since this in itself might generate valuable classroom discussion.

The main concern in developing diagnostic items is to ensure that the incorrect answer choices are connected to incorrect or incomplete understandings; and thus all student answers, not just the correct answers, provide insights into their learning.

CONCLUSIONS

Diagnostic items are not only a teaching resource but could also be used for professional development purposes. Pedagogical content knowledge, "that specialized amalgam of content and pedagogy that is uniquely the province of the teacher" (Shulman, 1987, p. 8), develops with experience. This knowledge connects content with teaching strategies, including multiple ways to present knowledge, and understanding how novice learners might address a topic differently from experts. Therefore, these diagnostic items could be used as discussion starting points for study groups. Novice teachers might be surprised at the ways in which students approach certain topics, and all teachers can gain by sharing strategies for what to do once certain misconceptions have been identified. Teachers could then work together to develop a variety of content-relevant and age-appropriate strategies to help students learn.

It is becoming more common for teachers to be presented with test data in the hope that they will use the information to

improve instruction. However, there is not always an immediate or obvious link between the test data and a teacher's daily instructional practice. In this project the issue is being approached from the opposite direction—starting with questions that teachers ask in their everyday practice for the purpose of finding out “where is this class right now?” and “what misconceptions are getting in the way of learning?” In other words, the goal is to help teachers better utilize questioning and discussion to improve student learning, a key strategy within formative assessment. By starting with the instructional decisions that teachers make, the intention is to create both a resource and a habit of mind that is of obvious value to teachers. Lesson planning is a natural part of every teacher's daily activities. Embedding a focus on questioning strategies—through carefully constructed diagnostic items—supports teachers in improving instruction by providing them with a natural extension to their practice, rather than requiring a major shift in mind-set.

Acknowledgement

We gratefully acknowledge the support of the Institute of Education Science without which this work would not have been possible. Funding provided under grant number R305K040051.

References

- Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory* (rev. ed.). Prospect Heights, IL: Waveland Press.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139-144.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham, UK: Open University Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1-22.

Cite This Article as: *Teachers College Record*, Date Published: January 10, 2006
<http://www.tcrecord.org> ID Number: 12285, Date Accessed: 3/15/2006 8:51:50 PM